UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Deep Learning Aided Visual Localisation in Urban Pedestrian Environments

by Maleen Jayasuriya

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of
Prof. Gamini Dissanayake and Dr. Ravindra Ranasinghe

University of Technology Sydney
Faculty of Engineering and Information Technology

July 2021

Certificate of Original Authorship

I, Maleen Jayasuriya declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program

Signed: Production Note: Signature removed prior to publication.

Date: 26/07/2021

Deep Learning Aided Visual Localisation in Urban Pedestrian Environments

by

Maleen Jayasuriya

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Abstract

Localisation of a mobile robot is a fundamental problem in robotics research. In a known environment, localisation can be performed using a prebuilt map, whereas much more complex simultaneous localisation and mapping (SLAM), which estimates both the robot location and the map, is required when operating in an unknown environment. This thesis focuses on the localisation of low-speed vehicles ranging from personal mobility devices to delivery robots, operating in a known outdoor urban environment using low-cost cameras with the objective of improving their functionality and safety. Existing techniques for vision only localisation, even while operating in known environments, requires SLAM due to the difficulty in building reliable maps that are persistent across long time frames.

This thesis proposes an approach that circumvents this problem by utilising convolutional neural network (CNN) based perception of (a) persistent pole-like landmarks such as lamp posts, trees, street signs and parking meters, and (b) important ground surface boundaries related to persistent infrastructure such as curbs, pavement edges and manhole covers, found in urban environments. Localisation is carried out on a prebuilt map consisting of the 2D locations of these landmarks and a vector distance transform (VDT) representation of the ground surface boundaries. An extended Kalman filter (EKF) fuses these observations to carry out pose estimation while robustly dealing with missed detections and wrong classifications.

This approach is further extended by utilising an omnidirectional camera to improve the effective field of view (FoV) of the landmark detection system. The framework utilises an information theoretic strategy to decide the best viewpoint to serve as an input to the CNN in a given iteration, instead of the full 360° coverage offered by an omnidirectional camera, in order to leverage the advantage of having a higher field of view without compromising on performance.

The final contribution of this thesis is a strategy to incorporate the knowledge of traversable paths in the environment into the overall localisation framework, as the target applications predominantly travel on pavements or footpaths. It is demonstrated that enforcing the constraint that the vehicle can only traverse on specific regions in a given map significantly enhances the quality of the location estimate. A decision making framework to ascertain whether the traversability constraint should be enforced or whether there has been a deliberate action to move the vehicle out of the predefined path is also presented.

Real-world experiments carried out in dynamic urban environments across large time gaps in the year and at different distance scales, utilising an instrumented mobility scooter, are presented to highlight the effectiveness of the contributions of this thesis in contrast to state of the art visual SLAM based approaches.

Acknowledgements

First and foremost, I would like to thank my supervisors, Professor Gamini Dissanayake and Dr. Ravindra Ranasinghe for taking me on as their student. Their advice, support and guidance has not only been instrumental in this work, but the example they set with their character, humility and work ethic will remain a constant inspiration.

I would like to express my gratitude to Dr. Janindu Arukgoda who was a constant source of advice and support both in terms of the theoretical as well as implementation aspects of this thesis. I would also like to thank Mr. Josh Olsen and Mr. Nathanael Gandhi whose support was instrumental in the development of the mobility scooter hardware platform. Likewise, the immense help received from Mr. Julien Collart, Mr. Peter Morris, Mr. David Murphy, and Ms. Kavindie Katuwandeniya in conducting experiments is greatly appreciated.

I would like to acknowledge the warmth and friendship of the staff and students at the University of Technology (UTS) Robotics Institute (RI) which made studying and working an absolute joy and privilege. The knowledge sharing sessions, reading groups, talks, and seminars in particular have been a great source of knowledge and inspiration. I am deeply appreciative of the many friendships that made the less glamorous aspects of my research, tolerable. This includes those who have been already mentioned, as well as Dr. Masha Popović, Mr. James Unicomb, Dr. Richardo Khonasty, Ms. Sheila Sutjipto, Mr. Yujun Lai, Mr. Stefano Aldini, Mr. Stefan Kiss, Mr Hongkyoon Byun, Ms. Mitchell Usayiwevu, Mr. Jesse Mehami, Ms. Lan Wu, Ms. Anna Lidfors Lindqvist, and many more who made my time at UTS RI a positive one.

I am also grateful to Dr. Godaliyadda and Dr. Parakarama, my supervisors at the University of Peradeniya for their guidance, support and particularly the advice to follow a career in academia.

Finally I would like to acknowledge the incredible support system made up of my family and friends in Sri Lanka. My brother and sister-in-law for taking care of me in Sydney (including the home cooked Sri Lankan meals). My sister for her support. My fiancée for being a constant rock, creative partner, and friend though the good times and bad. Last but not least to my parents for their immeasurable hard work, love, guidance and support. This thesis is dedicated to them.

Contents

List of Figures

List of Tables

Acronyms & Abbreviations

1D	One-Dimensional
2D	Two-Dimensional
3D	Three-Dimensional
UTS	University of Technology Sydney
RI	Robotics Institute
PMD	Personal Mobility Device
HMI	Human-Machine Interface
GNSS	Global Navigation Satellite Systems
LIDAR	Light Detection and Ranging
RADAR	Radio Detection and Ranging
SONAR	Sound Navigation and Ranging
SLAM	Simultaneous Localisation and Mapping
CNN	Convolutional Neural Network
EKF	Extended Kalman Filter
DOF	Degrees of Freedom
GPS	Global Positioning System
IMU	Inertial Measurement Unit

RTK-GNSS Real-time Kinematics GNSS

RTK-GNSS Differential GNSS

RTK-GNSS Assisted GNSS

RGB-D Red, Green, Blue and Depth

BA Bundle Adjustment

RTAB Real-Time Appearance-Based Mapping

HED Holistically Nested Edge Detection

YOLO You Only Look Once

SDT Signed Distance Transform

UDT Unsigned Distance Transform

VDT Vector Distance Transform

VO Visual Odometry

OGM Occupancy Grid Map

FoV Field of View

UV-Loc Urban Visual Localiser

RMSE Root Mean Square Error

Nomenclature

	General Notations
t	Time (continuous)
k	Time (discrete step)
υ	Linear velocity
ω	Angular velocity
U_t	A vector containing the linear and angular velocity at time t
θ_t^i	i^{th} bearing measurement to landmark observation at time t
Θ_t	A vector containing all bearing measurement to landmark observation at time t
l_t^i	i^{th} semantic label to landmark observation at time t
L_t	A vector containing all semantic labels of landmark observations at time t
M_L	Environmental landmark map
$(x_{z,t}^i, y_{z,t}^i)$	i^{th} ground surface boundary point at time t
Z_t	A vector ground surface boundary points at time t
DT_v	Vector Distance Transform
DT_x	x component of the Vector Distance Transform
DT_y	y component of the Vector Distance Transform
M_G	Ground surface boundary map

$\hat{\boldsymbol{X}}_t$	Pose estimate of robot in 2D space at time t. Consists of position components $\hat{x}_t$, $\hat{y}_t$ and the orientation component $\hat{\phi}_t$.
P_t	Robot pose covariance matrix at time t
$\square_{t-1}$	Previous state
$\square_{t\|t-1}$	Predicted current state
$\square_t$	Updated current state
$g(\cdot,\cdot)$	Motion model
$h(\cdot,\cdot)$	Observation model
$\nabla G_{\square}$	Jacobian of the motion model with respect to $\square$
$\nabla H_{\square}$	Jacobian of the observation model with respect to $\square$
ν	Innovation vector
S	Innovation covariance
K	Kalman gain
$\square_l$	EKF Update parameters related to landmark observations
$\square_g$	EKF Update parameters related to ground surface observations
vp	A set of perspective camera viewpoints obtained from an omnidirectional camera
ΔP^{vp}	A derived information metric that quantifies the impact on the overall pose uncertainty from a camera viewpoint
${}^{A}\boldsymbol{T}_{B}$	The transform matrix of frame B relative to frame A
$\square_u$	Refers to unconstrained states, covariances and related parameters
$\square_c$	Refers to constrained states, covariances and related parameters
${}^{nt}\square$	Pose components in local tangent frame
M_P	Polyline boundary map
ub_t	Upper path boundary at time t
lb_t	Lower path boundary at time t
d^2	Mahalanobis distance

$cov(\cdot)$	Covariance matrix
$diag(\cdot)$	Diagonal matrix
$trace(\cdot)$	Trace of a matrix
$\Sigma(\cdot)$	Summation
$\int$	Integral operator
∂	Partial differentiation operator